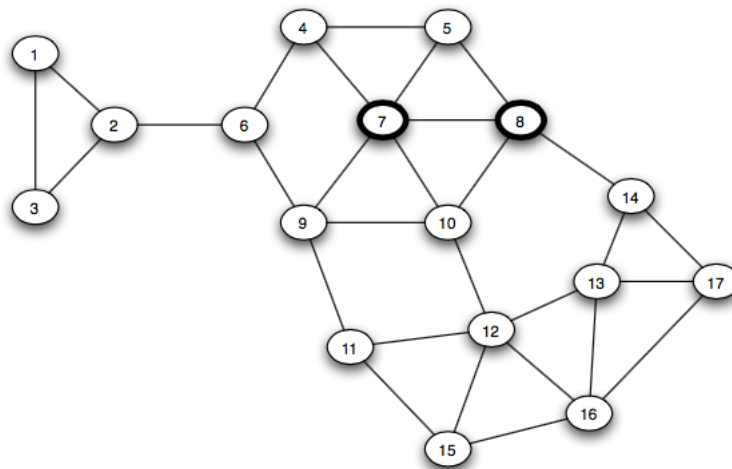


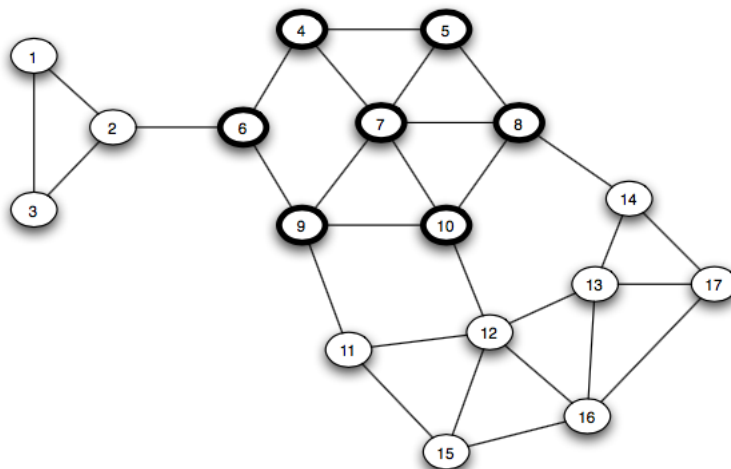
1 A Larger Example of Cascading Behavior

Consider what happens in the cascading behavior model on the following network with $a = 3, b = 2$, and the whole network starting with strategy B except for nodes 7 and 8:



(a) Two nodes are the initial adopters

We have that in this cascading behavior process with threshold $q = \frac{2}{5}$, A will spread to nodes $\{5, 10\}$ in step $t = 1$, nodes $\{4, 9\}$ when $t = 2$, and nodes $\{6\}$ when $t = 3$ but will not reach any more nodes in periods $t \geq 3$:



(b) The process ends after three steps

We can see that nodes $\{1, 2, 3\}$ and nodes $\{11, 12, 13, 14, 15, 16, 17\}$ both form clusters with density $\frac{2}{3}$, so a cascading process with threshold $q = \frac{2}{5}$ cannot penetrate these clusters since $\frac{2}{3} > 1 - q = \frac{3}{5}$.

2 Introduction to Epidemics

The way in which communicable diseases spread through groups of people is determined not only by the biological properties of the pathogen — such as its inherent contagiousness and severity — but also by the network structures within a population. In general, we call the relevant network structure a *contact network*, where there are nodes for each person and an edge exists between nodes if those two people come into contact with each other in a way that makes it possible for the disease to be spread from one to the other.

Note that the contagion and the network are very closely intertwined; even within the same population, different pathogens with different modes of transmission can have very different contact networks. For example, a highly contagious disease that spreads via airborne transmission will involve a huge number of edges and will be very connected, whereas something like an STD will be much sparser.

Now that we see that modeling the underlying contact network is critical to understanding the spread of an epidemic, we will turn to a basic model for the epidemic to better understand the spreading process.

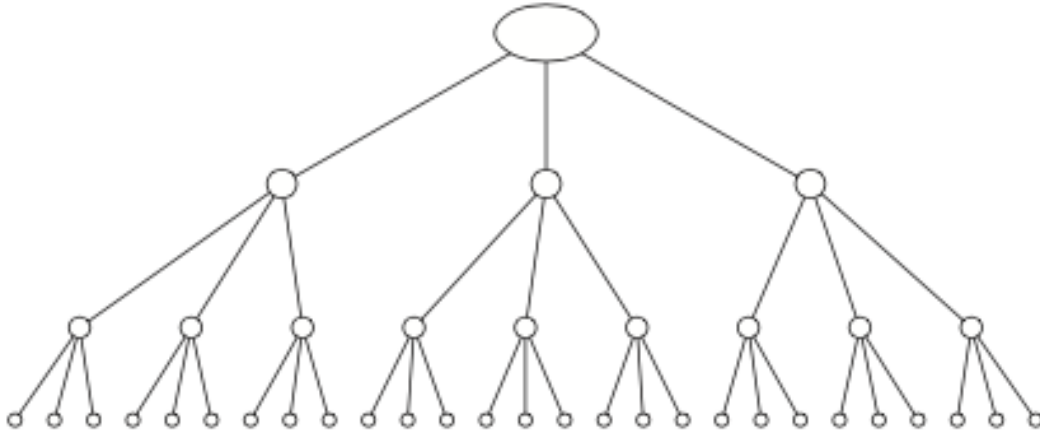
3 Branching Processes

One model for epidemics that we will study is called the *branching process*. Branching processes can be described as follows:¹

- *The first wave.* Suppose that a person carrying a new disease enters a population and transmits it to each person he meets independently with probability p . Furthermore, suppose that he meets k people while he is contagious; we call these k people the *first wave* of the epidemic. Based on the random transmission of the disease from the initial person, some of the people in the first wave may get infected with the disease, while others may not.
- *The second wave.* Now, each of the k people in the first wave goes out into the population and meets k more people (assumed to be distinct from the first k people), resulting in a second wave of $k \times k = k^2$ people. Each infected person in the first wave passes the disease independently to each of the k second-wave people they meet, again independently with probability p .
- *Subsequent waves.* Further waves are formed in the same way, by having each person in the current wave meet k new people and pass the disease to each independently with probability p .

An illustration of this branching process is below.

¹This description was adapted from Easley & Kleinberg, Ch21, Section 2.



Remember that each edge does not necessarily mean the disease was transmitted, but only that the two people came into contact; there is still probably p that the disease was transmitted for each edge; to better see this, below are two examples of the same branching process: one for a highly contagious pathogen (that is, one with a high p), and one with a low probability of spreading:



Notice that in the slow-spreading epidemic on the right, the branching process dies out after only two waves. So it is possible for the process to die out after some finite number of waves; this presents two questions:

First, can we ever be sure that the process will have died out by a given round? We can compute the probability that the process is still alive after n rounds to be $(1 - p)^n$. Therefore, we cannot be sure that the process will have died out by the n th wave (for some finite n), as the probability $(1 - p)^n$ is always greater than zero.

Second, can we ever be sure that the process will eventually die out? To answer this, we must examine the expected number of new cases for each individual; if this expectation is smaller than 1, then intuitively the process isn't able to "replenish" itself in each new wave, becoming smaller and smaller (in expectation) in each wave. If the expectation is greater than 1, then the number of infected is trending upwards as the number of waves increases, so the process could go on without ever dying out. This quantity, the expected number of new cases of the disease caused by a single individual, is called the *basic reproductive number* of the disease, and is denoted R_0 . We will discuss this quantity more next.

4 Basic Reproductive Number and a Dichotomy of Branching Processes

Recall from the previous section that R_0 is the *basic reproductive number*, or the expected number of new cases caused by a single individual. Since each person makes contact with k additional people

and infects each of them independently with probability p , we can see that $R_0 = pk$. Moreover, let q_n denote the probability that the epidemic survives at least n waves, and let $q^* = \lim_{n \rightarrow \infty} q_n$ be the probability that the epidemic goes on infinitely. This gives us the notation necessary to give Theorem 1:

Theorem 1. (a) $R_0 < 1 \implies q^* = 0$, and (b) $R_0 > 1 \implies q^* > 0$.

In other words, when the basic reproductive number is less than 1, the probability that the process never dies out is 0; that is, it is guaranteed to die out. When the basic reproductive number is greater than 1, however, there is a non-zero probability that the branching process goes on infinitely. How do we go about proving this theorem²? First, define X_n to be the number of infected individuals at level n . This brings us to the following lemma:

Lemma 2. $E(X_n) = R_0^n$.

To see this, note that there are k^n different people at level n (examine the example trees above to convince yourself of this) and the probability of each being infected is p^n , as the disease must make its way down the tree n levels for an individual at level n to be infected. Thus, $E(X_n) = k^n p^n = (pk)^n = R_0^n$.

How do we use this lemma to prove the theorem? It turns out this lemma can be used to directly show part (a) of the theorem; part(b) will require a little more work. Let us begin part (a) with the case when $R_0 < 1$.

Corollary 3. $R_0 < 1 \implies q^* = 0$

Consider that

$$\begin{aligned} E(X_n) &= \Pr[X_n = 1] + 2\Pr[X_n = 2] + 3\Pr[X_n = 3] + \dots \\ E(X_n) &= \Pr[X_n \geq 1] + \Pr[X_n \geq 2] + \Pr[X_n \geq 3] + \dots \end{aligned}$$

This is just a reformulation of $E(X_n)$ which is valid for any nonnegative X_n . From the second line, we can easily see that $E(X_n) \geq \Pr[X_n \geq 1]$. Now, we begin to notice that $\Pr[X_n \geq 1]$ is precisely the definition of q_n . Moreover, we know that $\lim_{n \rightarrow \infty} E(X_n) = \lim_{n \rightarrow \infty} R_0^n = 0$ when $R_0 < 1$. Since $E(X_n) \geq q_n$ and $\lim_{n \rightarrow \infty} E(X_n) = 0$, we must also have that $\lim_{n \rightarrow \infty} q_n = q^* = 0$, which concludes the proof.

Remember, this does not necessarily imply part (b). For example, suppose X_n is defined to be 4^n with probability 2^{-n} and 0 otherwise. Then, we have that $E(X_n) = 4^n \times 2^{-n} = 2^n$, which tends to infinity, whereas $\Pr(X_n > 0) = 2^{-n}$, which tends to 0. How, then, do we show part (b), namely, that $R_0 > 1 \implies q^* > 0$?

Lemma 4. $q_n = 1 - (1 - pq_{n-1})^k$

For a detailed explanation of this recursive formula for q_n , see Easley & Kleinberg, Section 21.8, pages 593-594. How can we use this formula to prove part (b) of the theorem? We can simply look at the sequence q_0, q_1, q_2, \dots and look at the limiting value of the sequence; to simplify things, we define $f(x) = 1 - (1 - px)^k$, so that the sequence we are interested in is

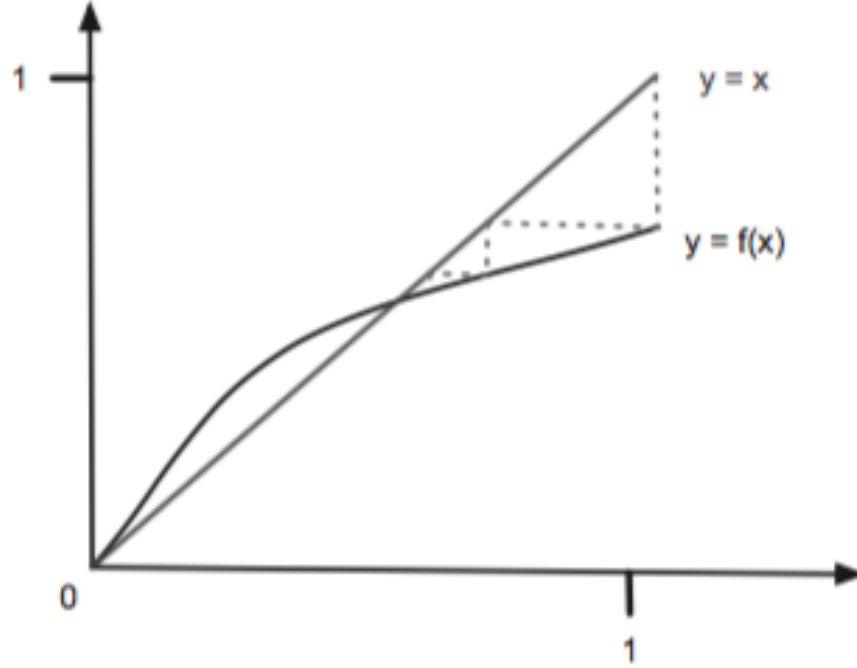
$$q_0, q_1, q_2, \dots = 1, f(1), f(f(1)), f(f(f(1))), \dots$$

To see the limiting behavior³ of this sequence, we look at the graph of $y = f(x)$ alongside $y = x$. If we begin at $(1, 1)$ on $y = x$ and we wish to get to the point $(f(x), f(x))$, we move first vertically to

²For a more detailed explanation of the proof, see Easley & Kleinberg, Ch21, Section 8, Part A.

³For an algebraic explanation of the proof rather than a graphical one, see Easley & Kleinberg page 595.

the $y = f(x)$ so that we are at $(x, f(x))$, and then move horizontally to the $y = x$ curve so that we are at $(f(x), f(x))$. If we continue this process indefinitely, we will eventually stop when the $y = x$ and $y = f(x)$ curves meet; this point of intersection, (x^*, x^*) , is the convergence of the sequence q_0, q_1, q_2, \dots and thus we can conclude that $q^* > 0$, as the point (x^*, x^*) lies on $y = x$ and is in the interval strictly between 0 and 1. A graphical explanation of this process is shown below.



You may be wondering, where in this proof did we use the assumption that $R_0 > 1$? In the graphic above, note that the derivative of $y = f(x)$, that is, $f'(x)$ is greater than 1 at $x = 0$ and then becomes less than 1 at some point before $x = 1$. The fact that $f'(0) > 1$ is critical to the proof and come from the fact that $f'(x) = pk(1 - px)^k \implies f'(0) = pk = R_0$ and the assumption that $R_0 > 1$. Therefore, we have shown that $R_0 > 1 \implies q^* > 0$.